

## Module 5

Data Source : Excel spreadsheet

Software : Pentaho Data Integration Software and Oracle Database.

Steps:

1.Requirements : First Excel file is loaded

Checked if any data is null or not and then added some validations.  
By doing some validations , checking the data type and constraints.  
Sort and Merge Operations are applied.

Step 1:

The screenshot shows the 'Filter rows' step configuration. The 'Step name' is 'Filter rows'. The 'Send 'true' data to step:' is 'Select values'. The 'Send 'false' data to step:' is empty. The 'The condition:' field contains a complex SQL WHERE clause:

```
BranchPlantKey IS NOT NULL
AND (
    Date IS NOT NULL
    AND (
        ItemMasterKey IS NOT NULL
        AND (
            TransTypeKey IS NOT NULL
            AND (
                CustVendorKey IS NOT NULL
                AND (
                    UnitCost IS NOT NULL
                    AND (
                        Currency IS NOT NULL
                        AND Quantity IS NOT NULL
                    )
                )
            )
        )
    )
)
```

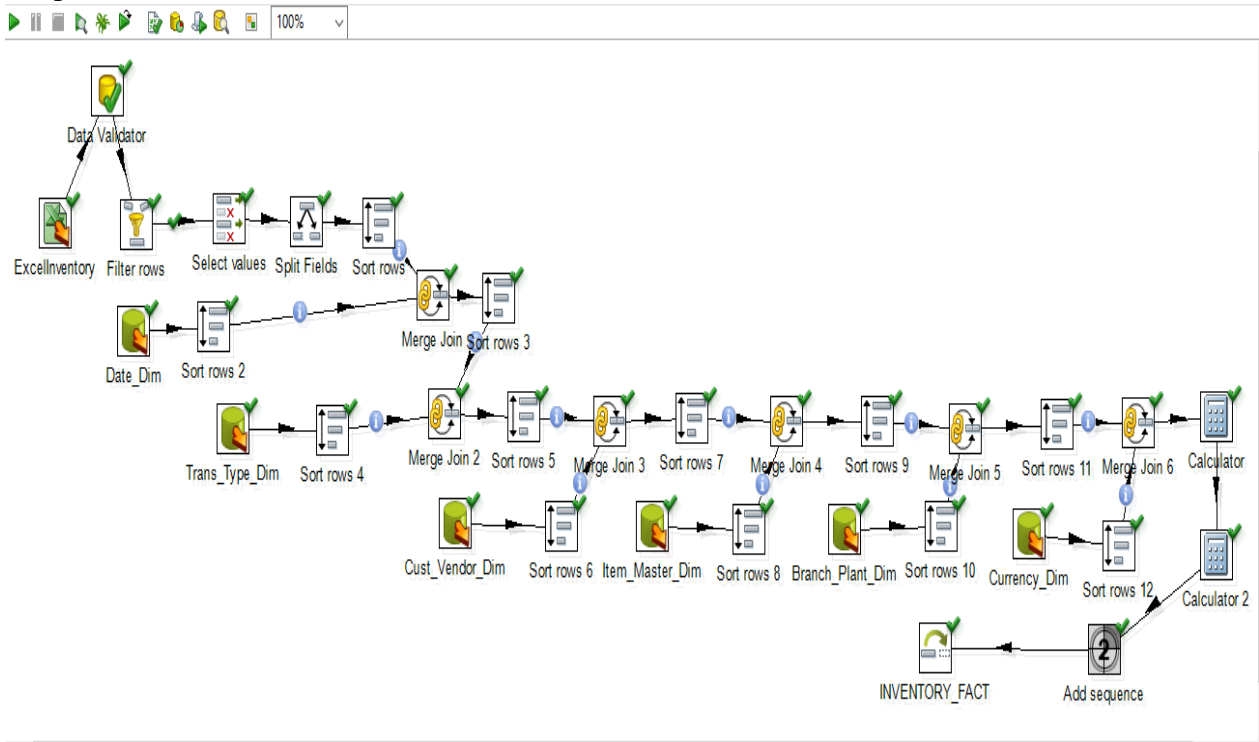
Step 2:

The screenshot shows the 'Data Validator' step configuration. The 'Stepname:' is 'Data Validator'. The 'Select a validation to edit:' list includes 'BranchPlantKey', 'Date', 'ItemMasterKey', 'TransTypeKey', 'CustVendorKey', 'UnitCost', 'Quantity', and 'Currency'. The 'BranchPlantKey' validation is selected. The configuration details are:

- ☐ Report all errors, not only the first
- ☐ Output one row, concatenate errors with separator : |
- Validation description: BranchPlantKey
- Name of field to validate: BranchPlantKey
- Error code: (empty)
- Error description: (empty)
- Type: (empty)
- Verify data type? ☒
- Data type: Integer
- Conversion mask: (empty)
- Decimal Symbol: (empty)
- Grouping Symbol: (empty)

### 3.Transformation:

#### Step 2:



1. Calculator1 => represents  $\text{Unit\_Cost(Inserting)} = \text{Unit\_Cost} * \text{Exchange\_Rate}$

2. Calculator2 => represents  $\text{Ext\_cost} = \text{Unit\_Cost(calculated from Calculator 1)} * \text{Quantity}$

Output: 10 rows will be affected.

#### Execution Results

Execution History / Logging / Step Metrics / Performance Graph / Metrics / Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	
9	Sort rows 4	0	5	5	0	0	0	0	0	Finished	0.0s	250	
10	Sort rows 3	0	11	11	0	0	0	0	0	Finished	1.0s	11	
11	Merge Join 2	0	16	11	0	0	0	0	0	Finished	1.3s	13	
12	Sort rows 5	0	11	11	0	0	0	0	0	Finished	1.3s	9	
13	Cust_Vendor_Dim	0	0	20	20	0	0	0	0	Finished	0.6s	32	
14	Sort rows 6	0	20	20	0	0	0	0	0	Finished	0.6s	32	
15	Merge Join 3	0	31	10	0	0	0	0	0	Finished	1.6s	20	
16	Sort rows 7	0	10	10	0	0	0	0	0	Finished	1.6s	6	
17	Item_Master_Dim	0	0	20	20	0	0	0	0	Finished	0.6s	32	
18	Sort rows 8	0	20	20	0	0	0	0	0	Finished	0.6s	32	
19	Merge Join 4	0	30	10	0	0	0	0	0	Finished	1.9s	16	
20	Branch_Plant_Dim	0	0	20	20	0	0	0	0	Finished	0.6s	32	
21	Sort rows 10	0	20	20	0	0	0	0	0	Finished	0.6s	32	
22	Sort rows 9	0	10	10	0	0	0	0	0	Finished	1.9s	5	
23	Merge Join 5	0	30	10	0	0	0	0	0	Finished	2.2s	14	
24	Currency_Dim	0	0	4	4	0	0	0	0	Finished	0.6s	6	
25	Sort rows 12	0	4	4	0	0	0	0	0	Finished	0.3s	14	
26	Sort rows 11	0	10	10	0	0	0	0	0	Finished	2.2s	4	
27	Merge Join 6	0	14	10	0	0	0	0	0	Finished	2.5s	6	
28	Calculator	0	10	10	0	0	0	0	0	Finished	2.5s	4	
29	Calculator 2	0	10	10	0	0	0	0	0	Finished	2.5s	4	
30	ExcellInventory	0	0	15	15	0	0	0	0	Finished	0.6s	24	
31	Data Validator	0	15	15	0	0	0	0	0	Finished	0.6s	24	
32	Add sequence	0	10	10	0	0	0	0	0	Finished	2.5s	4	
33	INVENTORY_FACT	0	10	10	10	10	0	0	0	Finished	2.5s	4	

Step 4: Loading ,we can see there are 10 new rows, that are get inserted in the database.

Script Output x

 Task completed in 2.565 seconds

INVENTORYKEY	BRANCHPLANTKEY	DATEKEY	ITEMMASTERKEY	TRANSTYPEKEY	CUSTVENDORKEY	UNITCOST	QUANTITY	EXTCOST
18310	11	89	5	3	15	11.03	20	220.6
18309	9	96	5	5	11	33.21	18	597.78
18308	7	95	3	1	13	21.22	15	318.3
18307	5	86	9	5	1	33.11	30	993.3
18306	1	84	19	1	5	11.22	10	112.2
18305	17	92	5	3	5	3.302	15	49.53
18304	7	87	7	1	3	3.322	40	132.88
18303	5	94	1	3	15	1.711	17	29.09
18302	3	85	17	3	7	2.233	20	44.66
18301	9	88	15	1	11	33.8954	50	1694.77